Multi-sensor Aerosol Products Sampling System

M. Petrenko[1,2], C. Ichoku[2], G. Leptoukh[2]

[1]Earth System Science Interdisciplinary Center, University of Maryland, College Park, Maryland

[2]NASA Goddard Space Flight Center, Greenbelt, Maryland

Aerosols are tiny particles suspended in the air that can be made up of sand, dust, smoke, smog, and so on. Over the glove, aerosols have a major impact on the air quality, hydrological cycle, and climate; still, the scale of this impact is still poorly understood. To fill this gap in our knowledge, global and local properties of atmospheric aerosols have been extensively observed and measured using both satellite and ground-based instruments, especially during the last decade. Aerosol properties retrieved by the different instruments contribute to an unprecedented availability of the most complete set of complimentary aerosol measurements ever acquired. However, some of these measurements remain underutilized, largely due to the complexities involved in analyzing them synergistically.

To characterize the inconsistencies and bridge the gap that exists between the sensors, we have established a Multi-sensor Aerosol Products Sampling System (MAPSS), which consistently and uniformly samples aerosol products from multiple spaceborne sensors, including MODIS (on Terra and Aqua), MISR, OMI, POLDER, CALIOP, and SeaWiFS. Samples of satellite aerosol products are extracted over Aerosol Robotic Network (AERONET) locations as well as over other locations of interest such as those with available ground-based aerosol observations.

In this way, MAPSS enables aerosol scientists across the world to compare and integrate data between aerosol observations from multiple sensors, enhancing our understanding of aerosols, and improving the quality of the aerosol measurements. In our paper, we explain the sampling methodology and concepts used in MAPSS, and demonstrate specific examples of using MAPSS for an integrated analysis of multiple aerosol products.

Multi-sensor Aerosol Products Sampling System

M. Petrenko[1,2], C. Ichoku[2], G. Leptoukh [2]

[1]{Earth System Science Interdisciplinary Center, University of Maryland, College Park, Maryland}

[2]{NASA Goddard Space Flight Center, Greenbelt, Maryland}

Correspondence to: M. Petrenko (maksym.petrenko@nasa.gov)

Abstract

Global and local properties of atmospheric aerosols have been extensively observed and measured using both spaceborne and ground-based instruments, especially during the last decade. Unique properties retrieved by the different instruments contribute to an unprecedented availability of the most complete set of complimentary aerosol measurements ever acquired. However, some of these measurements remain underutilized, largely due to the complexities involved in analyzing them synergistically. To characterize the inconsistencies and bridge the gap that exists between the sensors, we have established a Multi-sensor Aerosol Products Sampling System (MAPSS), which consistently samples and generates the spatial statistics (mean, standard deviation, direction and rate of spatial variation, and spatial correlation coefficient) of aerosol products from multiple spaceborne sensors, including MODIS (on Terra and Aqua), MISR, OMI, POLDER, CALIOP, and SeaWiFS. Samples of satellite aerosol products are extracted over Aerosol Robotic Network (AERONET) locations as well as over other locations of interest such as those with available ground-based aerosol observations. In this way, MAPSS enables a direct cross-characterization and data integration between Level-2 aerosol observations from multiple sensors. In addition, the available well-characterized co-located ground-based data provides the basis for the integrated validation of these products. This paper explains the sampling methodology and concepts used in MAPSS, and demonstrates specific examples of using MAPSS for an integrated analysis of multiple aerosol products.

1 Introduction

Atmospheric aerosol parameters are routinely retrieved and archived for public access by an array of ground-based and spaceborne sensors, especially since the 1990s (Holben et al. 1992, 1998; Herman et al., 1997; Husar et al., 1997; Deuzé et al., 1999; Mishchenko et al., 1999; Ignatov and Stowe 2002; Chu et al., 2002; Remer et al., 2002; Hsu et al., 2004; Ichoku et al., 2005; Torres et al., 2007; Winker et al., 2007; Levy et al., 2009). However, the integrated use of these observations is greatly complicated by the numerous discrepancies and differences that exist between the sensors and their aerosol products, including dissimilar spatial and temporal resolutions, archival strategies, approaches to quality control, and so forth (Kahn et al., 2007; Liu and Mishchenko 2008; Li et al., 2009). The problem is further complicated by the differences in the algorithms, underlying assumptions, and uncertainties involved in creating the products.

The purpose of this paper is to introduce the Multi-sensor Aerosol Products Sampling System (MAPSS), which is a framework designed to provide a uniform and consistent sampling of aerosol products from multiple sources. MAPSS was originally designed to support the validation effort for the aerosol retrieval algorithm from the Moderate-resolution Imaging Spectro-radiometer (MODIS) sensor aboard the Terra and Aqua satellites (Ichoku et al., 2002), but has been redesigned with the aim of facilitating an integrated use and detailed comparative analysis of aerosol measurements from multiple satellite sensors. Currently, the supported sensors include MODIS on Terra and Aqua, the Multi-angle Imaging Spectro-Radiometer (MISR) on Terra, the Ozone Monitoring Instrument (OMI) on Aura, the POLarization and Directionality of the Earth's Reflectances (POLDER) on Parasol and its heritage ADEOS and ADEOS-2 satellites, the Cloud-Aerosol Lidar with Orthogonal Polarization (CALIOP) on Calipso, as well as aerosol retrievals using the Deep Blue algorithm from the Sea-viewing Wide Field-of-view Sensor (SeaWiFS) aboard the SeaStar spacecraft. Like the original MAPSS, this multi-sensor version is also based on the collocation of the satellite data products over the global AERosol Robotic NETwork (AERONET) of ground-based sun-photometer stations and over other important sites.

The relevant characteristics of the aerosol data products from the different sensors are described in Sect. 2, while the details of the MAPSS sampling concepts are explained in Sect. 3. The implementation of MAPSS on the Web, online data access and analysis approaches,

and user tips are described in Sect. 4. Section 5 discusses possible applications of the proposed system, followed by conclusions in Sect. 6.

2 Supported aerosol products

This work focuses on aerosol observations from multiple ground-based and spaceborne instruments; the supported instruments, products, and some of their key characteristics are outlined in Table 1. It is pertinent to note that all satellite data products supported in MAPSS are derived directly from the retrieval level aerosol products (Level 2); the Level 2 data represents the highest available spatial resolution for each product/sensor combination and is free of aggregation artifacts that can be present in data at Level 3 (Levy et al., 2009; Zhang and Reid, 2010; Hyer et al., 2011). The remainder of this section provides a brief description of each of the products supported in MAPSS, while highlighting unique aerosol properties reported in these products.

AERONET (http://aeronet.gsfc.nasa.gov) sun-photometers measure aerosol properties using ground-based observations of solar direct and diffuse radiances in three observation configurations (direct solar, principal plane, and almucantar), based upon which they provide three distinct aerosol product categories. The first is the aerosol optical depth or thickness (AOD or AOT) product, which is obtained from the AERONET direct measurements of solar irradiance. The second is the spectral deconvolution aerosol product (SDA), which is a more advanced product that uses a spectral deconvolution algorithm to derive additional aerosol properties, not available by means of the direct retrievals. Finally, the third is the inversion aerosol product (INV), which is an aerosol product that uses an inversion algorithm that, based on a limited set of measured aerosol properties, estimates possible values for other properties. Additionally, for the convenience of aerosol data inter-comparison and validation, MAPSS provides an auxiliary dataset with AERONET AOD interpolated to the common wavelengths used in the spaceborne retrievals based on the established wavelength dependence of AOD (Eck et al., 1999). However, it is important to note that the interpolated dataset is not quality-assured by the AERONET team and may contain inaccuracies that are inherent in the process of interpolation.

The MODIS (http://modis.gsfc.nasa.gov) aerosol product (MOD04 and MYD04) comprises the ambient aerosol optical thickness and other physical properties of aerosols retrieved globally over land and ocean. It should be noted that the dataset name prefix 'MOD' signifies

the MODIS instrument onboard the Terra satellite, while 'MYD' indicates the MODIS instrument onboard the Aqua satellite. All the MODIS aerosol products currently included in MAPSS are those retrieved at 10-km nominal resolution at nadir. In addition to the aerosol product, MAPSS also includes the MODIS precipitable water vapor product designated by MOD05 and MYD05 for Terra and Aqua, respectively. The MODIS precipitable water vapor product is of two kinds, one based on infrared (IR) retrieval at 1-km nominal spatial resolution, and the other from near-infrared retrieval at 5-km nominal spatial resolution.

The MISR (http://www-misr.jpl.nasa.gov) aerosol product (MIL2ASAE) features aerosol retrievals based on observations from 9 independent camera angles. Multiple viewing angles allow MISR to measure certain aerosol properties that are not available from the other instruments (e.g., aerosol particle size). Furthermore, MISR multiple cameras enable retrievals under conditions that are unfavorable to single-view (e.g. nadir) instruments, such as over bright surfaces or sun glint, where the other instruments are unable to make reliable retrievals in the visible wavelengths.

The OMI (http://www.knmi.nl/omi/research/instrument/index.php) aerosol product (OMAERUV) measures the near-UV (near ultraviolet) aerosol absorption and extinction optical depth, as well as single scattering albedo, among other aerosol properties. Moreover, OMI is capable of retrieving absorption optical depth in partially cloudy conditions that usually pose a challenge to other aerosol instruments.

The POLDER (http://www.icare.univ-lille1.fr/parasol) aerosol land product (PxL2TLGC) and aerosol ocean product (PxL2TOGC) are derived from measuring spectral, directional, and polarized properties of reflected solar radiation. One of the main features of the POLDER instrument is its utilization of polarization properties of the measured radiation for retrieving anthropogenic aerosol optical depth.

The CALIOP (http://www-calipso.larc.nasa.gov) aerosol product (05kmALay) represents atmospheric curtain slices portraying the vertical distribution of aerosols and clouds in the atmosphere, including the density and certain properties of individual aerosol layers.

The SeaWiFS (http://disc.sci.gsfc.nasa.gov/dust/) aerosol product (SWDB) uses the Deep Blue (Hsu et al., 2004) algorithm to derive aerosol optical thickness and Ångström exponent. The key features of this product are the retrievals of aerosol properties over both bright desert and vegetated surfaces, and a highly precise calibration of the SeaWiFS sensor.

Since each of the foregoing data sets has a few versions because of the periodic revisions and updates of their retrieval algorithms over time, the data versions that were current at the time of MAPSS development were processed. Collection 051 was processed for both Terra and Aqua MODIS, Version 0022 for MISR, Version 003 for OMI, Version 3-01 for CALIOP, Version K for POLDER, Version 002 for SeaWiFS (preliminary version), Version 2 for the AERONET AOD and INV products, and Version 4.1 for the AERONET SDA product. Therefore, unless otherwise specified, all of the illustrations and analyses shown in this paper are based on these data versions for the respective satellite sensors, and Version 2 for AERONET.

3 Data Sampling and Analysis

The proposed framework expands on the concepts of the sampling approach that was developed by Ichoku et al. (2002), and used for validation and analysis of MODIS aerosol products using AERONET measurements (Remer et al., 2002; Chu et al., 2002; Ichoku et al., 2003, 2005). In the original approach, the spatial aerosol measurements acquired aboard the Terra and Aqua spacecraft were sampled within 50x50 km areas, centered over AERONET sun photometer measurement sites, as well as over certain other point locations where the satellite aerosol data samples are required. The pixel containing the ground station was determined by finding a pixel with the minimal Euclidian distance between the longitude/latitude coordinates of the center of this pixel and those of the ground station. Next, the extent of the sampling area was determined by finding surrounding pixels located no more than 25 km from the central pixel, based on the Euclidean distance between their longitude and latitude coordinates. In turn, temporal measurements from each ground-based location were sampled at each satellite overpass time in 1-hour sampling segments (i.e., 30 minutes before and 30 minutes after the overpass).

The Multi-sensor Aerosol Products Sampling System (MAPSS) described in this paper uses a similar sampling approach and includes data sampling from a wider variety of spaceborne aerosol sensors, as outlined in Table 1. To accommodate the various resolutions of the extended set of the supported aerosol products and to make their ground sampled areas as equivalent as possible, instead of sampling within the nominal 50x50-km square used for MODIS in the original MAPSS, the multi-sensor data sampling space is now defined by a circle of approximately 50-km diameter that is centered on the ground (e.g., AERONET)

measurement station (see Figure 1). However, while testing the 50-km diameter sampling, it was found that, because of the wide variety of pixel shapes and sizes among the satellite products, some of the coarser-resolution products were not sampled in a balanced way around the ground site, even at nadir. For instance, if POLDER data (nominal resolution: 19 x 19 km at nadir) were sampled within a 50-km diameter circle, with the ground point located slightly off the center of a pixel, only the neighbors of the central pixel on the same side as the ground point would be sampled, while those on the opposite side would be omitted. Therefore, based on empirical analysis of the different data products, it was found that a diameter of 55 km would enable overall balanced sampling within the circular sample space for the different data products, at least near nadir. In effect, a pixel is sampled only if the distance between its center and the ground station does not exceed 27.5 km, calculated from the longitude/latitude coordinates of the ground station and the pixel's center using the Haversine distance formula (Sinnott 1984). The actual number of possible pixels within this 55-km diameter sample space depends on the pixel shape and size for each sensor. Also, for a given sensor, since the pixel size typically increases in size away from nadir, the maximum number of pixels within the sample space decreases with the distance of the ground station away from the nadir of the satellite scene. Given their respective nominal spatial resolutions (see Table 1), the maximum number of pixels within the 55-km diameter sample space at nadir for the different sensors are as follows: MODIS - 25, MISR - 9, OMI - 8, POLDER - 9, CALIOP - 11, and SeaWiFS - 16.

To determine the effects (if any) of using the circular (instead of square) sample space on the sample statistics, means of Terra MODIS Collection 5 AOD derived from both sample spaces (i.e. 50 x 50-km square and 55-km diameter circle) were calculated and plotted against corresponding statistics from AERONET for comparison (Figure 2). The results indicate that the difference in the shape of the sampling space has only a small effect on the derived sample statistics of the data. However, it was found that the circle-based sampling produced approximately 22% fewer data points: 2,394,624 data points were generated for the studied 10-year period using the square-based sampling, while only 1,881,858 data points were produced using the circle-based sampling. This difference can be explained by recalling that, when the ground station is located off-nadir of the sensor or off-center of the central pixel in the sample space, in order to maintain a uniform sampling area, the number of pixels in a circle-based sample can be much reduced even down to 1, depending on the sensor observation geometry and how far the ground station is located off-nadir. This can result in a null subset, if none of the reduced number of sample points contains aerosol retrieval. In

contrast, the number of pixels in the square-based samples remains roughly constant regardless of the retrieval geometry conditions, resulting in the larger number of data points. Additionally, the further inspection of the data revealed that, while the number of the data points decreased uniformly over most of the sampling locations, it increased in certain locations, including islands, coastal areas, and locations in high latitudes. This difference can be attributed to the greater accuracy of the Haversine formula used in the new circle-based sampling compared to the Euclidean distance approach that was used previously. In the MAPSS application, the Haversine formula is accurate within approximately 200 meters, which allows a high precision for improving sampling in the above-mentioned areas, where aerosol retrievals are scarcer.

3.1 Statistics of Sampled Data

Data products with very different spatial resolutions have been sampled within the uniform 55-km diameter circular sample space in order to render them as comparable as possible using appropriate analysis tools. As such, the data subsets sampled over each ground station during each satellite overpass time are used to calculate a number of appropriate statistics, including their mean, median, mode, and standard deviation. Corresponding statistics are also calculated from the temporal subsets of the ground-based measurements. In either case, only data points with valid values of each measured parameter are used to calculate the statistics. The ranges of valid data values are established a-priori based on the "valid_range" pair of values specified in the metadata for some of the data products. Where such valid range is not specified in the data, a quantitatively reasonable valid range is assumed. For instance, 0° – 180° is set as the valid range for 'solar zenith angle' because this is the known normal range for this type of zenith-to-nadir angular parameter. By specifying the valid ranges for different parameters, fill values (usually large negative numbers like -9999) are used in the different data sets to indicate non-retrieval or other data gaps, as well as other spurious values that may occur in any of the data products, which are automatically excluded from the statistics calculations. No other filtering is applied to the data before the statistics are calculated. The statistics include:

- **ndat** - total number of data points sampled (e.g., number of pixels contained within the 55-km sample space in a satellite dataset);

- **nval** - number of sampled data points with valid data for use in the statistics calculation;
- **cval** - value of a central sampling data point in the subset. For the spaceborne data, this is a point in the sample that has the smallest distance to the ground station. For the ground-based data, this is a point in the sample that is the closest in time to the overpass of the satellite;
- **mean** - mean of the subset;
- **medn** - median of the subset;
- **mode** - mode of the subset;
- **sdev** - sample standard deviation of the subset. (Note: when **nval**≤1, this value is undefined as opposed to zero).

It is important to note that **mean**, **median**, and **sdev** are computed only for continuous datasets (i.e., datasets comprising real numbers). On the other hand, **mode** is computed only for discrete datasets (i.e., datasets comprising only integer numbers, for example, data quality flags).

It is pertinent to note that, since the ground-based AERONET measurements are sampled and recorded in the MAPSS archive for every satellite overpass, in cases where multiple satellites pass over a particular location within a 1-hour timeframe (e.g., satellites in the A-Train formation), a single AERONET measurement can be sampled and recorded in the archive multiple times. It is recommended to account for this duplication when further aggregation of the AERONET data in the MAPSS archive is performed, in order to avoid possible oversampling issues.

3.2 Characterizing the spatio-temporal variability of the data

The sampling of the satellite aerosol data beyond the pixels lying directly over the ground stations is intended to provide not only the average values of the measured parameters, but also their local variability and other characteristics (over the 55-km sample space around the station) that can enhance detailed scientific research and validation. Therefore, in addition to computing the basic statistical parameters, a linear multiple regression plane is fitted to each continuous spatial subset, and a linear regression line to each continuous temporal subset (Fox 1997; Ichoku et al., 2002). Based on this fit, the following additional statistics are computed:

- **slop** - slope of the fitted plane or line;

- **slaz** - azimuth (direction) of the slope of the plane;
- **mcoc / lcoc** - multiple correlation coefficient (for a spatial subset regressed on the lat/lon coordinates of the sample points) or linear correlation coefficient (for a temporal subset regressed on the measurement times of the ground-based data samples such as those of AERONET).

These (**slop**, **slaz**, and **mcoc/lcoc**) statistics are not computed unless the number of sampled valid data points is sufficient to obtain a statistically robust fit for the plane or line. To ensure that this condition is met, the shape, size and maximum number of pixels that fall within the nadir sampling area of each sensor were carefully examined. Based on this empirical analysis, the minimal required number of valid data points to fit a plane was determined for each sensor so that these points do not form a degenerate plane; in other words, the data points in a subset should not all fall on the same line. In this way, the minimal numbers of valid data points required to fit a plane were set for the different aerosol products as: MODIS - 10, MISR - 5, OMI - 4, POLDER - 5, SeaWiFS - 7. Likewise, the minimal number of valid data points required to fit a line for AERONET and CALIOP aerosol products was set to 2, as this is the minimal number of points that can be used to uniquely define a line.

As shown in Figure 3 and Figure 4, these statistics can be used to assess the local spatio-temporal distribution and variation of the samples. In particular, the azimuth (i.e., **slaz**) parameter indicates the direction of the gradient of an aerosol parameter under consideration, pointing toward the lower values of the parameter. For instance, if AOD is the parameter of interest, **slaz** would typically indicate the direction of decreasing aerosol density, which would generally correspond with the direction of wind flow and plume dispersion from the aerosol source. Even when a sampling area contains multiple aerosol plumes, **slaz** for AOD can still point in the direction from the optically thickest to the optically thinnest plume, as demonstrated in Figure 4. In such cases, in addition to **slaz**, it is helpful to consider the spatial slope (**slop**) and multiple correlation coefficient (**mcoc**) also, as lower values of these spatial statistics would indicate a single homogeneous plume, while higher values might indicate multiple plumes, or a strong aerosol source present in the 55-km-diameter sample space.

3.3 Quality assurance (QA) data

The Level-2 aerosol products from all of the satellite sensors supported in MAPSS include quality assurance / quality control (QA) flags that indicate the "trustworthiness" of individual

pixels in these datasets. For MODIS and SeaWiFS, aerosol QA flags are integer numbers ranging from 0 to 3, with 3 representing the highest quality. For the MISR and OMI data, the reverse is the case (i.e., 0 is the highest quality). Finally, for POLDER and CALIOP, QA data are a combination of one or more flags, most of which are real numbers ranging between 0 and 1, where 1 indicates the highest quality. Table 2 provides a short summary of the discussed QA flags, while for detailed guidelines on the usage of these flags, depending on a particular science application, it is advised to consult the up-to-date science team recommendations for the analyzed products.

The QA values are set by the product retrieval algorithms and can be used to screen the data of pixels with potentially uncertain or erroneous retrievals or even invalid values. To facilitate such screening, MAPSS extracts the QA flags over the sampling area and computes the statistical mode for integer QA flags and mean for real QA flags. These statistical modes of the integer QA flags and means of the real QA flags provide an equivalent of a single number quality assessment for each sample, and can be used to screen the corresponding subset statistics, as demonstrated in Figure 5.

To test whether the approach of screening the already computed statistics of the aerosol parameters based on the statistically aggregated values of the QA flags (Ichoku et al., 2002, 2003) has the potential of being less effective than screening individual pixels using their respective QA flags before computing the sample statistics discussed above, the two approaches were compared as outlined in Table 3 and Table 4. It was found that the screening of the aerosol parameter subset statistics by their QA mode produces results that are similar to the screening of individual pixels by their QA before computing the statistics, although the former method results in slightly fewer data points, since entire data points are rejected based on the average QA value even if some of its component individual pixels have good QA flags. However, this is a small trade-off compared to the increased amount of effort involved in pre-screening before statistics.

4 Data Management and Accessibility

The subset statistics of hundreds of aerosol parameters and ancillary data sampled from daily measurements of six satellite sensors, over hundreds of ground-based stations, constitute an enormous amount of data, whose data management and accessibility requirements are non-trivial. Therefore, special tools and resources were developed to handle data batch processing,

storage, updates, and access, as seamlessly as possible. Data management is handled through a custom-designed database, while data access is through a Web interface.

4.1 Batch Processing

An automated software system has been developed to perform the multi-sensor aerosol sampling and statistical analysis in batch mode at pre-specified times. The MAPSS system interrogates the list of ground sites to determine the current list of sites over which data sampling is to be performed. It then fetches the aerosol data products from the online data sources of the supported satellite sensors to extract data subsets and derive their relevant statistics, as described in the previous section. Similarly, available AERONET data are obtained and sampled to derive corresponding temporal statistics. These analyses are performed on a daily basis, and the derived subset statistics are archived in simple comma-separated text (CSV) files that are easily accessible online (http://modis-atmos.gsfc.nasa.gov/MAPSS/; as of August 2011, this archive contained over 1,420,000 CSV files). However, in some cases, there may be a delay of several days between the time the data becomes available in a particular sensor's data repository and the time the extracted subset statistics become accessible in MAPSS. This delay is associated with the time required for data retrieval and processing. Also, there are likely to be some data gaps resulting from the inability of the different sensor algorithms to retrieve certain aerosol parameters due to unfavorable conditions, such as certain types of surface characteristics, clouds in reflective bands, sun glint over ocean, or even data downtime for sensor calibrations.

4.2 The MAPSS Database and Data Structure

The multi-sensor aerosol data analysis requires exploration of how to organize such a diverse set of data in a coherent, user-friendly, manner to facilitate access and analysis. While a basic examination of the data can be effectively performed using the described comma-separated text (CSV) file archive, the sheer amount of the sampled data and the limitations of the text file access make a more sophisticated analysis less practical. Therefore, it was necessary to establish a relational database that would streamline the access and the querying of the data. For this purpose, a dedicated PostgreSQL (http://www.postgresql.org) database was created, with the data structure designed to precisely reflect the logical organization of the sampled data.

The database is organized as a collection of individual data records, where each data record stores a sampled set of measurements that are acquired by a specific sensor at a specific location and at a specific time, and a corresponding list of ancillary data, including the measurement geometry (e.g., solar zenith at the time of the measurement, sensor azimuth, etc.) and QA information. Additionally, each data record is associated with data provenance information such as the name of the aerosol data file sampled, and the location of the central sampling point (**cval**) in this file. This information provides for quick access to the original data and allows for more detailed exploration of data points of interest.

Therefore, each data record consists of the statistics that are computed for one or more aerosol parameters retrieved by a specific aerosol sensor, as described in Sections 3.1 and 3.2. For example, AOD, fine mode fraction, reflectance, and other aerosol parameters in the Aqua MODIS aerosol product (MYD04), sampled at 18:08 UTC over the GSFC site on 2010-07-06, comprise a single data record, whereas **mean**, **sdev**, **mcoc,** and other statistics computed from the data sample for AOD at 550nm constitute a single statistics record.

The database is routinely populated with the most current information from the MAPSS CSV files. In August 2011, the database contained 28,818,432 data records and 1,613,178,051 statistics records. This size, however, poses a set of database maintenance challenges, where the operations necessary to keep the database consistent and up-to-date require a substantial computation time. Therefore, the information in the database can be less current than the information in the CSV archive.

It is expected that as new product versions become available, they will also be extracted and organized in the MAPSS database, such that it would also be possible, if desired, to access and compare different versions of aerosol retrievals from the same sensor.

4.3 Web MAPSS

A special point-based data analysis system was created within the framework of the Giovanni system (Acker and Leptoukh 2007; Berrick et al., 2009) to provide for a simple and customized Web-based access to the data archived in the MAPSS database (http://giovanni.gsfc.nasa.gov/mapss/). A screenshot of this so-called Web MAPSS data-access interface is shown in Figure 6. Through this interface, it is possible to select desired sampling locations and time period, aerosol products and associated data sets, as well as statistical variables of interest and their desired range of data quality (QA). Based on these

criteria, Web MAPSS selects appropriate MAPSS data samples, filters them according to the specified QA values, aligns the data in time and space, and produces a graphic plot of the data. The selected data sets are also formatted and staged for download as a comma-separated CSV text file. In this way, Web MAPSS allows the user to quickly and intuitively assess a combination of aerosol properties from multiple satellite sensors collocated with AERONET ground-based measurements over one or more ground locations, without having to download vast amounts of data from disparate product archives. In addition, MAPSS provides documentation explaining the basic attributes of each sampled product, complete with links to the relevant algorithm theoretical basis documents (ATBD) of the original products, thereby saving the time needed for locating these documents and facilitating an exploration of unfamiliar aerosol products.

5 Applications

Several possible applications of the data sampled by the MAPSS system have been envisioned. For example, the sampled data can be used for comparing spaceborne observations with corresponding ground-based measurements. Based on such comparison, it would be possible to assess the accuracy of aerosol retrievals from multiple spaceborne instruments in a manner similar to the validation studies of the MODIS aerosol products (Remer et al., 2002, 2005, 2008; Chu et al., 2002; Ichoku et al., 2003, 2005; Levy et al., 2010).

Simultaneous comparison of aerosol retrieval accuracy from multiple satellite sensors can help investigate the intrinsic strengths and weakness of the different instruments for aerosol remote sensing over different regions of the globe. For instance, appropriate comparative analysis results could provide indications of which sensors are particularly suitable for analysis of aerosols in a given region of the globe, as well as to explore the peculiarities of aerosol retrievals from a particular instrument over this region. Figure 8 shows an example, where measurements of aerosol optical depth (AOD) from multiple spaceborne sensors over two locations, (a) Evora (Portugal) and (b) Bondville (USA), are compared to the corresponding interpolated measurements from AERONET. The spaceborne sensors and retrieval algorithms used in the comparison include Terra MODIS Dark Target - land (TMOD DT), Terra MODIS Deep Blue - land (TMOD DB), Aqua MODIS Dark Target - land (AMOD DT), Aqua MODIS Deep Blue - land (AMOD DB), MISR, OMI, and CALIOP. A

comparison of the regression equations and squared correlation coefficients from the different sensors over the two locations highlight differences that exist in aerosol retrievals from multiple sensors.

Also, the spatial statistics collected by MAPSS can provide an insight into the spatial consistency of spaceborne retrievals. For example, Figure 9 shows linear regression fits of (a) **mean**, and (b) central values (i.e., **cval**) of AOD retrievals from the same set of satellite sensors against the corresponding interpolated AERONET data. The **mean** and **cval** plots show fairly similar fits, with the MODIS **cval** fits being slightly better, indicating that spaceborne MODIS AOD retrievals directly over Dakar (Senegal) are more accurate than the overall retrievals in the area surrounding Dakar. This could probably be explained by difficulties associated with retrieving aerosol properties over complex mixed environments due to their inherent surface inhomogeneities, particularly given the tendency for the occurrence of complex mix of marine aerosols, Saharan dust, urban pollution, and smoke from biomass burning in that region.

Another synergistic use of the collocated satellite and AERONET data is that the unique retrievals of aerosol layer heights provided by Calipso-CALIOP can be used to evaluate the degree of uncertainty due to the occurrence of multiple aerosol layers or variation in the layer heights of single or multiple layers in the data produced by the other sensors. In the example shown in Figure 10, AOD differences between ground-based observations from AERONET and spaceborne observations from both POLDER (ocean) and OMI over Dakar are higher in the presence of multiple aerosol layers, based on layer retrievals from Calipso-CALIOP.

Finally, the MAPSS database can be used as a data source for other aerosol investigation projects. For example, the AeroStat system (http://giovanni.gsfc.nasa.gov/aerostat/) uses the MAPSS database for assessing systematic biases that can possibly exist in aerosol measurements retrieved by different sensors.

Conclusions

The Multi-sensor Aerosol Products Sampling System (MAPSS) provides a consistent sampling approach that enables easy and direct inter-comparison and validation of the diverse aerosol products from different satellite sensors in a uniform and consistent way. The range of statistics collected in MAPSS facilitates the investigation of various spatio-temporal

properties of aerosols, as observed from multiple sensors with complementary capabilities, thereby helping to expand our understanding of the distribution and environmental impact of aerosols from different perspectives at local scales, with the possibility of extension by aggregation to global scales. Indeed, the readily available unified access to distinct aerosol parameters from multiple sensors provides a platform for acquiring a more complete understanding of the inter-relationships that may exist between the different physical properties of aerosols, which cannot all be measured from one or even a few sensors. It is expected that the MAPSS system will open the way for a multitude of synergistic aerosol studies, some of which have probably not been considered till date.

Acknowledgements

Support for the development of this project has been provided by NASA HQ under Grant Number NNX08AN39A through the ROSES 2007 ACCESS Program based on a proposal entitled: "Integrated validation, intercomparison, and analysis of aerosol products from multiple satellites". We thank the science and support teams of MODIS, MISR, OMI, POLDER, CALIOP, SeaWiFS and AERONET for retrieving and making available their respective aerosol products, as well as for providing assistance during the development of MAPSS sampling for these products. We also thank the Principal Investigators (PIs) of AERONET sites and their staff for establishing and maintaining these sites. In addition, we would like to thank the Giovanni team for developing and hosting the MAPSS database and the Web MAPSS user interface.

References

Acker, J., and Leptoukh, G.: Online Analysis Enhances Use of NASA Earth Science Data, EOS, Trans. Am. Geophys. Union, 88, 14, 2007.

Berrick, S., Leptoukh, G., Farley, J., and Rui, H.: Giovanni: A Web services workflow-based data visualization and analysis system, IEEE Trans. Geosci. Remote Sens., 47, 106 - 113, 2009.

Chu, D. A., Kaufman, Y. J., Ichoku, C., Remer, L. A., Tanre, D., and Holben, B. N.: Validation of MODIS aerosol optical depth retrieval over land, Geophys. Res. Lett., 29, 2002.

Deuzé, J. L., Herman, M., Goloub, P., Tanré, D., and Marchand, A.: Characterization of aerosols over ocean from POLDER/ADEOS-1, Geophys. Res. Lett., 26, 1421–1424, 1999.

Eck, T., Holben, B., Reid, J., Dubovik, O., Smirnov, A., O'Neill, N., Slutsker, I., and Kinne, S.: Wavelength dependence of the optical depth of biomass burning, urban, and desert dust aerosols, J. Geophys. Res., 104, 31333–31349, 1999.

Fox, J. : Applied Regression Analysis, Linear Models, and Related Methods. Thousand Oaks: SAGE Publications, 1997.

Herman, M., Deuzé, J. L., Devaux, C., Goloub, P., Bréon, F. M., and Tanré, D.: Remote sensing of aerosols over land surfaces including polarization measurements and application to POLDER measurements, J. Geophys. Res., 102, 17039-17049, 1997.

Holben, B. N., Vermote, E., Kaufman, Y. J., Tanre, D., and Kalb, V.: Aerosol retrieval over land from AVHRR data—application for atmospheric correction, IEEE Trans. Geosci. Remote Sens., 30, 212-222, 1992.

Holben, B. N., Eck, T. F., Slutsker, I., Tanre, D., Buis, J. P., Setzer, A., Vermote, E., Reagan, J. A., Kaufman, Y. J., Nakajima, T., Lavenu, F., Jankowiak, I., and Smirnov, A.: AERONET—A federated instrument network and data archive for aerosol characterization, Remote Sens. Environ., 66, 1–16, 1998.

Hsu, N. C., Tsay, S. C., King, M. D., and Herman, J. R.: Aerosol properties over bright-reflecting source regions, IEEE Trans. Geosci. Remote Sens., IEEE Trans. Geosci. Remote Sens., 42, 557–569, 2004.

Husar, R. B., Prospero, J. M., and Stowe, L. L.: Characterization of tropospheric aerosols over the oceans with the NOAA advanced very high resolution rediometer optical thickness operational product, J. Geophys. Res., 102, 16889–16910, 1997.

Hyer, E. J., Reid, J. S., and Zhang, J.: An over-land aerosol optical depth data set for data assimilation by filtering, correction, and aggregation of MODIS Collection 5 optical depth retrievals, Atmos. Meas. Tech., 4, 379-408, 2011.

Ichoku, C., Chu, D. A., S. Mattoo, Kaufman, Y. J., Remer, L. A., Tanre, D., Slutsker, I., and Holben, B. N.: A spatio-temporal approach for global validation and analysis of MODIS aerosol products, Geophys. Res. Lett., 29, 2002.

Ichoku, C., Remer, L. A., Kaufman, Y. J., Levy, R., Chu, D. A., Tanre, D., and Holben, B. N.: MODIS observation of aerosol and estimation of aerosol Radiative forcing over southern Africa during SAFARI 2000, J. Geophys. Res., 108, 2003.

Ichoku, C., Remer, L. A., and Eck, T. A.: Quantitative evaluation and intercomparison of morning and afternoon Moderate Resolution Imaging Spectroradiometer (MODIS) aerosol measurements from Terra and Aqua, J. Geophys. Res., 2005.

Ignatov, A., and Stowe, L.: Aerosol Retrievals from Individual AVHRR Channels. Part I: Retrieval Algorithm and Transition from Dave to 6S Radiative Transfer Model, J. Atmos. Sci., 59, 313–334, 2002.

Kahn, R. A., Garay, M. J., Nelson, D. L., Yau, K. K., Bull, M. A., Gaitley, B. J., Martonchik, J. V., and Levy, R. C.: Satellite-derived aerosol optical depth over dark water from MISR and MODIS: Comparisons with AERONET and implications for climatological studies, J. Geophys. Res., 112, D18205.1-D18205.25, 2007.

Levy, R. C., Leptoukh, G. G., Kahn, R., Zubko, V., Gopalan, A., and Remer, L. A.: A Critical Look at Deriving Monthly Aerosol Optical Depth From Satellite Data, IEEE Trans. Geosci. Remote Sens., 47, 2942–2956, 2009.

Levy, R. C., Remer, L. A., Kleidman, R. G., Mattoo, S., Ichoku, C., Kahn, R., and Eck, T. F.: Global evaluation of the Collection 5 MODIS dark-target aerosol products over land, Atmos. Chem. Phys., 10, 10399-10420, 2010.

Li, Z., Zhao, X., Kahn, R., Mishchenko, M., Remer, L., Lee, K.-H., Wang, M., Laszlo, I., Nakajima, T., and Maring, H.: Uncertainties in satellite remote sensing of aerosols and impact on monitoring its long-term trend: a review and perspective, Ann. Geophys., 27, 2755-2770, 2009.

Liu, L., and Mishchenko, M. I.: Toward unified satellite climatology of aerosol properties: direct comparisons of advanced level 2 aerosol products, J. Quant. Spectrosc. Radiat. Transfer, 109, 2376–2385, 2008.

Mishchenko, M. I., Geogdzhayev, I. V., Cairns, B., Rossow, W. B., and Lacis, A. A.: Aerosol retrievals over the ocean by use of channels 1 and 2 AVHRR data: sensitivity analysis and preliminary results, Appl. Opt., 38, 7325-7341, 1999.

Remer, L.A., Tanre, D., Kaufman, Y. J., Ichoku, C., Mattoo, S., Levy, R., Chu, D. A., Holben, B. N., Dubovik, O., Smirnov, A., Martins, J. V., Li, R. -R., and Ahmad, Z.: Validation of MODIS aerosol retrieval over ocean, Geophys. Res. Lett., 29, 2002.

Remer, L.A., Kaufman, Y. J., Tanre, D., Mattoo, S., Chu, D. A., Martins, J. V., Li, R. -R., Ichoku, C., Levy, R. C., Kleidman, R. G., Eck, T. F., Vermote, E., and Holben, B. N.: The MODIS aerosol algorithm, products and validation, J. Atmos. Sci., 62, 947-973, 2005.

Remer, L.A., Kleidman, R. G., Levy, R. C., Kaufman, Y. J., Tanre, D., Mattoo, S., Martins, J. V., Ichoku, C., Koren, I., Yu, H., and Holben, B. N.: Global aerosol climatology from the MODIS satellite sensors, J. Geophys. Res., 113, D14S07, 2008.

Sinnott, R. W.: Virtues of the Haversine, Sky and Telescope, 68, 159, 1984.

Torres, O., Tanskanen, A., Veihelmann, B., Ahn, C., Braak, R., Bhartia, P. K., Veefkind, P., and Levelt, P.: Aerosols and surface UV products from Ozone Monitoring Instrument observations: An overview, J. Geophys. Res., 112, D24S47, 2007.

Winker, D. M., Hunt, W. H., and McGill, M. J.: Initial performance assessment of CALIOP, Geophys. Res. Lett., 34, L19803, 2007.

Zhang, J., and Reid, J. S.: A decadal regional and global trend analysis of the aerosol optical depth using a data-assimilation grade over-water MODIS and Level 2 MISR aerosol products, Atmos. Chem. Phys., 10, 10949–10963, 2010.

Table 1. Atmospheric aerosol measurement instruments and products supported in MAPSS. The indicated equator crossing times are based on the original orbital designs, and can change during the lifetimes of the satellites

Sensor	**Platform**	**Product**	**Spatial Resolution**	**Equator crossing time**	**Data period**
AERONET	N/A	AOT, SDA, INV	N/A	N/A	Varies with sites
MODIS	Terra Aqua	MOD04, MOD05 MYD04, MYD05	10x10 km	10:30 am 1:30 pm	Jan'00- Jul'02-
MISR	Terra	MIL2ASAE	17.6x17.6 km	10:30 am	Jan'00-
OMI	Aura	OMAERUV	13.7x23.7 km	1:38 pm	Oct'04-
POLDER	ADEOS ADEOS-2 PARASOL	P[1-3]L2TLGC P[1-3]L2TOGC	19x19 km	1:30 pm	Oct'96-Jun'97 Apr'03-Oct'03 Mar'05-
CALIOP	CALIPSO	05kmALay	5x5 km	1:32 pm	Jun'06-
SeaWiFS	SeaStar	SWDB	13.5x13.5 km	12:00 pm	Jan'98-Dec'10

Table 2. Summary of the quality assurance / quality control (QA) flags and their values for the aerosol products supported in MAPSS.

Products	QA Flag (name in MAPSS)	Usage notes	Values
MOD04 MYD04	QA-l	Land datasets	0=No confidence 1=Marginal 2=Good 3=Very good
	QAavg-o	Ocean datasets	
	QAdpbl-l	Deep Blue datasets	
MIL2ASAE	QAb	AOD and Ångström Exponent (AExp)	0=Successful retrieval, single mixture 1=Successful retrieval, multiple mixtures 2=Data filled by averaging neighboring pixels 3=No retrieval
	QAprop	SSA, AOD fraction, number fraction, and volume fraction	0=Good 1=Bad
OMAERUV	QAfaf	AOD, Absorption AOD, and SSA	0=Most reliable 1=Reliable 2=Less reliable
P[1-3]L2TLGC P[1-3]L2TOGC	QAinv	All datasets	Real number between 0 (Bad) and 1 (Excellent)
	QAviewgeom		
	pixQbits-l	Land datasets	32-bit flag field
	pixQbits-o	Ocean datasets	
05kmALay	flagDay	Indicates daylight retrievals	0=No, 1=Yes
	flagStratFeature	Indicates detected stratospheric features	0=No, 1=Yes
	flagBaseExtended	Indicates that the algorithm increased the vertical extent of the ground layer	0=No, 1=Yes
	modeNLayers	Mode of number of layers over all sampled pixels	Number of layers: [0, 8]
	cNLayers	Number of layers over the central pixel	
	CADscoreX	Additional layer-specific flags indicating confidence in retrievals. X stands for layer index, X=[1,8].	Number: [-100,100]
	QAlayIABfctrX		Real number: [0,1]
	QA0532extX QA1064extX		16-bit flag field
SWDB	QAaod	AOD – land and ocean	0=No confidence 1=Marginal 2=Good 3=Very good
	QAaod-l	AOD – land	
	QAaod-o	AOD – ocean	
	QAaexp	AExp – land and ocean	
	QAaexp-l	AExp – land	
	QAaexp-o	AExp – ocean	

Table 3. Correlation coefficient (R) between **mean** values of AERONET AOD and Terra MODIS AOD at 550nm depending on different MODIS data screening scenarios, and also between **cval** (value of the central pixel in the sample) values of these two products. Factors considered for the screening include QA (Quality Assurance) flags of individual data pixels, mode of QA flags of all valid pixels in the sample, and nval (number of valid pixels in the sample). MODIS QA value of 3 indicates the data with the best quality. Correlation values for the **medians** of AOD are not shown as they are only marginally different from the reported values for the **means** of AOD. AERONET data was interpolated to 550nm and screened to have nval of at least 4.

R of AOD at 550nm MOD04, V.5.1		**nval**					
		all	**>=20% (>=5)**	**[1,5]**	**[6,10]**	**[11,20]**	**[21,26]**
mean							
Land and Ocean	No QA filtering	0.84	0.86	0.79	0.86	0.87	0.91
	QA mode=3	0.90	0.90	0.90	0.91	0.90	0.91
	QA pixel=3	0.88	0.91	0.85	0.90	0.92	0.93
Land	No QA filtering	0.82	0.86	0.75	0.85	0.87	0.91
	QA mode=3	0.90	0.90	0.88	0.90	0.90	0.91
	QA pixel=3	0.82	0.86	0.75	0.85	0.87	0.91
Ocean	No QA filtering	0.89	0.92	0.88	0.92	0.93	0.91
	QA mode=3	0.93	0.92	0.92	0.93	0.97	0.86
	QA pixel=3	0.89	0.92	0.88	0.92	0.93	0.91
Deep Blue	No QA filtering	0.70	0.69	0.74	0.75	0.60	0.73
	QA mode=3	0.79	0.79	0.80	0.82	0.77	0.76
	QA pixel=3	0.75	0.84	0.70	0.82	0.88	0.89
cval							
Land and Ocean	No QA filtering	0.83	0.84	0.82	0.84	0.82	0.86
	QA mode=3	0.88	0.88	0.87	0.90	0.86	0.87
	QA pixel=3	0.90	0.91	0.86	0.91	0.91	0.92
Land	No QA filtering	0.83	0.84	0.81	0.84	0.82	0.86
	QA mode=3	0.87	0.88	0.87	0.90	0.86	0.87
	QA pixel=3	0.83	0.84	0.82	0.84	0.82	0.86
Ocean	No QA filtering	0.93	0.92	0.92	0.93	0.93	0.97
	QA mode=3	0.94	0.94	0.94	0.94	0.97	N/A
	QA pixel=3	0.93	0.92	0.93	0.92	0.93	0.97
Deep Blue	No QA filtering	0.72	0.72	0.78	0.78	0.63	0.71
	QA mode=3	0.80	0.80	0.81	0.83	0.78	0.77
	QA pixel=3	0.85	0.86	0.82	0.84	0.89	0.90

Table 4. Correlation coefficient (R) between **means**, **medians**, and **cvals** (value of the central pixel in the sample) of AERONET AOD and OMI AOD at 500 nm depending on different OMI data screening scenarios. Factors considered for the screening include QA (Quality Assurance) flags of individual data pixels, mode of QA flags of all valid pixels in the sample, and nval (number of valid pixels in the sample). OMI QA value of 0 indicates the data with the best quality. AERONET data was interpolated to 500 nm and screened to have nval of at least 4. OMI data was additionally screened to exclude data points with AOD>=6, which comprise 130 of approximately 50,000 data points when computing the statistics for "No QA filtering", and 5 of approximately 35,000 data points when computing the statistics for "QA mode=0".

R of AOD at 500nm OMAERUV, V.3	**nval**			
	all	**>=20% (>=2)**	**[1,5]**	**[6,10]**
mean				
No QA filtering	0.41	0.43	0.44	0.42
QA mode=0	0.52	0.51	0.56	0.42
QA pixel=0	0.55	0.55	0.57	0.52
median				
No QA filtering	0.42	0.45	0.42	0.45
QA mode=0	0.56	0.56	0.60	0.49
QA pixel=0	0.55	0.57	0.56	0.53
cval				
No QA filtering	0.43	0.43	0.41	0.41
QA mode=0	0.51	0.52	0.57	0.41
QA pixel=0	0.54	0.57	0.56	0.46

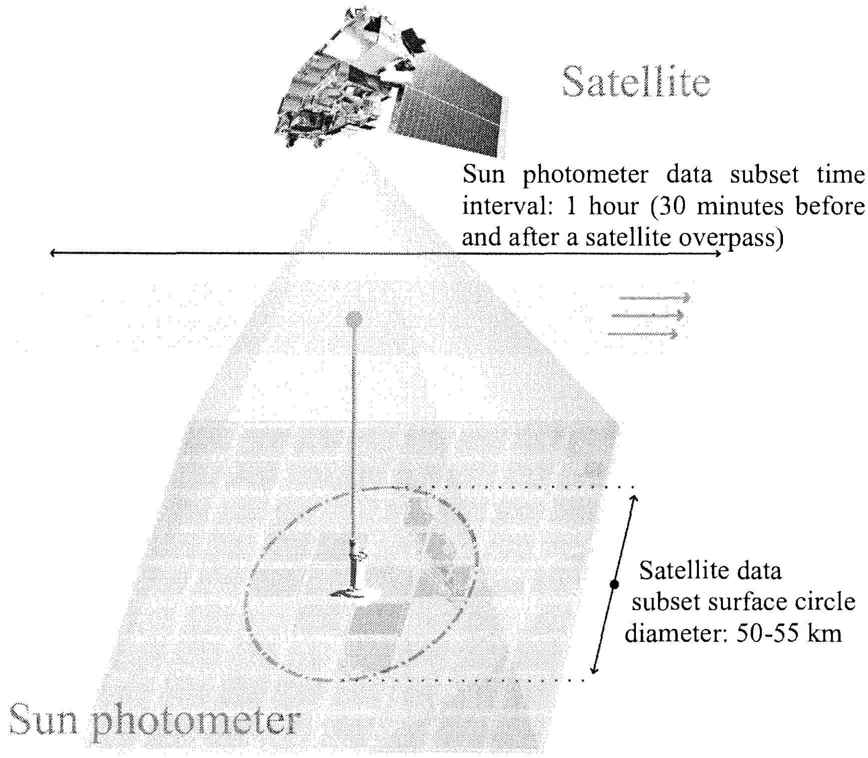

Figure 1. Overview of the sampling framework used in MAPSS. Sampling of each spatial spaceborne aerosol product involves extracting values of the pixels that fall within an approximate radius of 27.5 km from the chosen locations. Similarly, ground-based temporal observations in a particular location are sampled from measurements taken within 30 minutes before and 30 minutes after a satellite passes over this location.